PCT World Intellectual Property Organization
International Bureau

INTERNATIONAL APPLICATION PUBLISHED IN ACCORDANCE WITH THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶: **C07H 21/00**	**A2**	**(11) International Publication Number:** **WO 99/47536** **(43) International Publication Date:** September 23, 1999 (9/23/99)

(21) International Application Number: PCT/DE99/00755

(22) International Filing Date: March 18, 1999 (3/18/99)

(30) Priority Data:
198 12 103.2 March 19, 1998 (3/19/98) DE

(71) Applicant and Inventor: BERNAUER, Annette [DE/DE]; Weberstrasse 38, D-79249 Merzhausen (DE).

(81) Designated States: AU, BR, CA, CN, CZ, IL, JP, KR, MX, PL, RU, SG, UA, US; European Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
Without International Search Report; to be republished upon receipt of report.

(54) Title: METHOD FOR SYNTHESIZING NUCLEIC ACID MOLECULES

(57) Abstract:

The invention relates to a method for synthesizing nucleic acid molecules. The invention relates in particular to such a method which is carried out in a recursive manner. The nucleic acid constituents are preferably of synthetic or semisynthetic origin. The principle of the method according to the invention is based on attaching and/or linking an additional nucleic acid molecule to a prepared nucleic acid molecule, masking the end of the prepared nucleic acid molecule if no additional nucleic acid molecule is attached or linked thereto, and cleaving the additional nucleic acid molecule at a predetermined location, preferably resulting in an end to which an additional nucleic acid molecule may be attached and/or linked. The aforementioned method steps are optionally repeated as often as necessary until the desired product is synthesized. The invention further relates to a kit for carrying out the method according to the invention.

Method for synthesizing nucleic acid molecules

The invention relates to a method for synthesizing nucleic acid molecules. The invention relates in particular to such a method which is carried out in a recursive manner. The nucleic acid constituents are preferably of synthetic or semisynthetic origin. The principle of the method according to the invention is based on attaching and/or linking an additional nucleic acid molecule to a prepared nucleic acid molecule, masking the end of the prepared nucleic acid molecule if no additional nucleic acid molecule is attached or linked thereto, and cleaving the additional nucleic acid molecule at a predetermined location, preferably resulting in an end to which an additional nucleic acid molecule may be attached and/or linked, and the aforementioned method steps are optionally repeated as often as necessary until the desired product is synthesized. The invention further relates to a kit for carrying out the method according to the invention.

In the past 20 years, recombinant techniques for manipulating nucleic acids have given a tremendous impetus to many scientific disciplines, including the pharmaceutical industry and the field of medical research. In many fields of application it is desirable to prepare a nucleic acid molecule having a precisely defined sequence, in the simplest manner possible and with little expenditure of time and money. Currently, the most commonly used methods for preparing such nucleic acid molecules involve the cloning of DNA, for example from cDNA gene banks, optionally combined with subsequent sequencing of the isolated cDNA. On the other hand, DNA having a desired sequence may be prepared synthetically using the conventional phosphoramidite method.

Common methods for preparing desired double-stranded nucleic acid molecules are explained below, using the preparation of DNA molecules as an example. DNA molecules of interest must be isolated, for example by cDNA cloning or positional cloning, and cloned into suitable vectors. The propagation of the resulting vectors, and thus of the DNA molecules of interest, occurs in vivo. For this purpose, the vectors must be introduced into suitable host cells, for example bacteria or yeasts. For further manipulation of the DNA, for example for preparing modified constructs which mediate new phenotypic properties, the DNA must be isolated from the host organisms. Only then is it available for purposes of manipulation. For further propagation the DNA must be introduced once again into suitable host organisms. Thus, a large number of method steps and/or complicated manipulations are often necessary to prepare desired DNA. Unfortunately, as known to one skilled in the art, this level of effort may be several times greater if a larger quantity of various types of DNA is to be prepared.

Another method known in the prior art for in vitro synthesis of double-stranded DNA is the PCR technique. The availability of suitable matrix DNA is a prerequisite for such preparation. The subcloning of suitable DNA fragments and the possibly laborious adjustment of the correct reaction conditions for the PCR may delay the experimental work considerably.

The above-described methods known in the prior art are still relatively time-consuming and therefore

cost-intensive. Furthermore, as in the case of cDNA cloning, they are not always successful. The synthetic generation of longer nucleic acid fragments often poses significant difficulties in practice. Although the generation of DNA by PCR has greatly advanced DNA recombination technology, in some cases it may not be successful or may encounter problems, as described above.

The object of the present invention, therefore, is to provide a method which allows nucleic acid molecules of desired sequence and length to be easily synthesized in a time-saving manner. This object is achieved by the embodiments characterized in the claims.

The invention therefore relates to a method for synthesizing nucleic acid molecules, including part or all of the following steps:

1. Preparing a nucleic acid molecule that has at least one end which allows an attachment and/or linkage to an additional nucleic acid molecule;
2. Attaching and/or linking at least one additional nucleic acid molecule to the nucleic acid molecule, wherein one end of the at least one additional nucleic acid molecule is attached and/or linked to the at least one end of the nucleic acid molecule, and for the case of a linkage, the other end of the at least one additional nucleic acid molecule is masked;
3. Optionally masking the at least one end of the nucleic acid molecule which has not been attached and/or linked to an additional nucleic acid molecule;
4. Cleaving the at least one additional attached and/or linked nucleic acid molecule at a predetermined location, wherein the masking is removed, and one end is generated which allows an attachment and/or linkage to an additional nucleic acid molecule; and
5. Repeating steps (2) through (4) at least once, optionally multiple times, wherein in each case suitable nucleic acid molecules are used in step (2).

In one preferred embodiment of the method according to the invention, the additional nucleic acid molecule is a single-stranded nucleic acid molecule.

In one particularly preferred embodiment, after step (2) the method according to the invention includes the following step:
(2a) Replenishing the second nucleic acid strand, which has a sequence that is complementary to the single strand, using a polymerase activity, the masking optionally being removed beforehand.

In another particularly preferred embodiment, after step (4) or (5) the method according to the invention includes the following step:
(4/5a) Replenishing the second nucleic acid strand, which has a sequence that is complementary to the single strand, using a polymerase activity.

As mentioned above, the method according to the invention is suitable for synthesizing single-stranded (dsDNA) or partially double-stranded DNA.

The principle of the method according to the invention is illustrated in Figure 1. Further embodiments are illustrated in Figures 2 through 7.

In one embodiment of the method according to the invention, a single-stranded nucleic acid molecule, a partially double-stranded nucleic acid molecule having an overhanging 5' or 3' end, or a double-stranded nucleic acid molecule having a smooth end is prepared. In the next step, a single-stranded nucleic acid molecule is covalently bonded to this end of the prepared nucleic acid molecule using a ligase activity, for example a T4 RNA ligase. The single-stranded nucleic acid molecule may be bound at its 5'-phosphate or 3'-hydroxy end to the prepared nucleic acid molecule. Accordingly, the method according to the invention includes a nucleic acid synthesis in the 3' to 5' direction or in the 5' to 3' direction (based on the orientation of the precursor molecules). In these embodiments it is essential that the end of the single-stranded nucleic acid molecule which is not linked to the prepared nucleic acid molecule is masked. Within the meaning of the present invention, "masked" means that in this ligation assay, this end cannot be linked to an additional single-stranded nucleic acid molecule of the same type, resulting in single-stranded molecules that are composed of multiple copies of the same nucleic acid molecule, and which may likewise be linked to the prepared nucleic acid molecule. Within the meaning of the invention, masking is a chemical, enzymatic, or some other modification of the end which prevents the above-described linkage. Maskings within the meaning of the present invention are described in greater detail below. After the ligation, the ends of the prepared nucleic acid molecules which have not been linked to a single-stranded nucleic acid molecule are masked. In the next step, the single-stranded nucleic acid molecule which has been ligated to the prepared nucleic acid molecule is cleaved at a predetermined location, wherein the masking is removed and an end is generated which allows a linkage with a next single-stranded nucleic acid molecule. By means of the last two steps, the method according to the invention advantageously ensures that in further ligation steps only those nucleic acid molecules to which a single-stranded nucleic acid molecule has been ligated in the preceding step are further extended. The ligation, masking, and cleavage steps may be repeated in sequence as often as desired, in each case using molecules to be newly attached, and using suitable single-stranded nucleic acid molecules in each case.

In another preferred embodiment of the method according to the invention, after synthesis of the complete desired single strand the counterstrand having a complementary sequence is synthesized, using a polymerase activity. If the single strand has been synthesized in the 3' to 5' direction, and a double-stranded nucleic acid molecule having a smooth end or a partially double-stranded nucleic acid molecule has been prepared, the complementary nucleic acid strand may be synthesized directly from the free 3' end of the prepared nucleic acid molecule. However, if a single-stranded nucleic acid molecule has been prepared, and the synthesis of the single strand is carried out in the 3' to 5' direction, prior to the polymerase reaction a 3' end must be provided to the prepared nucleic acid molecule by hybridization of a suitable single-stranded nucleic acid oligomer. If the synthesis of the nucleic acid single strand has been carried out in the 5' to 3' direction, the last single-stranded nucleic acid molecule

that is ligated to the synthesized nucleic acid single strand is advantageously selected in such a way that the 3' end forms a hairpin structure, so that after cleavage of the hairpin structure a 3' end is provided for synthesizing the complementary nucleic acid strand, using a polymerase activity.

In another preferred embodiment of the method according to the invention, directly after (each) cleavage of the ligated single-stranded nucleic acid molecule and before the ligation of the next single-stranded nucleic acid molecule the corresponding complementary nucleic acid strand is synthesized using a polymerase activity. Otherwise, essentially the same procedure is followed as described above for synthesizing the complete complementary nucleic acid strand. In the case of a 5' to 3' synthesis direction, each single-stranded nucleic acid molecule is selected in such a way that it advantageously forms a hairpin structure at the 3' end. In addition, the masking at the 3' end of the ligated nucleic acid molecule is removed prior to the polymerase reaction which is carried out before the ends of the prepared nucleic acid molecules are masked.

In another embodiment of the method according to the invention, the additional nucleic acid molecules which are attached and/or thus linked to the prepared nucleic acid molecule are double-stranded. In this embodiment the prepared nucleic acid molecule is single-stranded or partially double-stranded, with an overhanging 3' or 5' end. If the additional nucleic acid molecule has a corresponding overhanging 3' or 5' end with a complementary sequence, attachment takes place by hybridization of the single-stranded overhanging ends. The other end of the additional nucleic acid molecule is preferably smooth. The above-described masking ensures that at this end no attachment to cohesive ends of additional nucleic acid molecules of the same type takes place via hybridization.

In yet another embodiment of the method according to the invention, the additional nucleic acid molecule is single-stranded, and attachment to the prepared nucleic acid molecule takes place via hybridization of complementary end-position nucleotides. In this embodiment the prepared nucleic acid molecule is single-stranded or partially double-stranded, with an overhanging 3' or 5' end. The additional single-stranded nucleic acid molecule may optionally also be covalently linked to the prepared nucleic acid molecule, using a ligase activity. If the hybridization occurs via 3' end-position nucleotides, in the next step the complementary strand may be synthesized using a polymerase activity. If the hybridization occurs via 5' end-position nucleotides, the 3' end of the additional single-stranded nucleic acid molecule is selected in such a way that it forms a hairpin structure, so that a 3' end is provided for the subsequent polymerization reaction for synthesizing the complementary nucleic acid strand. In the next step the synthesized nucleic acid double strand is cleaved at a predetermined location, and the recognition sequence which is necessary for the cleavage as well as the smooth end, i.e., the hairpin structure, is removed, resulting in a preferably cohesive end which allows attachment via hybridization and optionally a covalent linkage of the nucleic acid molecule to an additional single-stranded nucleic acid molecule.

The present invention also encompasses methods whose attachment, masking, and/or cleavage steps represent combinations of the corresponding steps of the above-mentioned embodiments. Thus, for

example, in a first synthesis cycle a single-stranded nucleic acid molecule may be covalently linked to a prepared nucleic acid molecule, the complementary nucleic acid strand may be subsequently synthesized, the double strand may be cleaved as described above, and in the next synthesis cycle another single-stranded nucleic acid molecule may be attached by hybridization.

If single-stranded nucleic acid molecules are linked to a prepared nucleic acid molecule using a ligase activity, it is not necessary to carry out the synthesis of the complementary nucleic acid strand after each attachment, masking, and/or cleavage step, or at the end of the synthesis of the complete nucleic acid single strand. The time at which the complementary strand is replenished may be selected as desired, in the sense that it is selected, for example, after a given permuting attachment, masking, and/or cleavage step.

Within the meaning of the present invention, the term "masking" means that a covalent linkage of two nucleic acid molecules is not possible. Masked single-stranded 3' ends may be generated, for example, by introducing an amino block, a dideoxynucleotide, a 3'-phosphate, or by using a synthetically introduced 5' end. Within the meaning of the present invention, masked single-stranded 5' ends are characterized, for example, by a missing phosphate group or by the introduction of a 5'-modified nucleotide (biotin-dNTP, digoxygenin-dNTP, for example). If an extension of a prepared nucleic acid molecule is carried out by hybridization of complementary end-position nucleotides, a double-stranded nucleic acid molecule having a smooth end at which an additional nucleic acid molecule is not able to hybridize at its end-position nucleotides is also referred to as "masked" within the meaning of the present invention. Thus, a partially double-stranded nucleic acid molecule having overhanging single-stranded ends may be masked by removing an overhanging 3' end, using an exonuclease activity, or by synthesizing the complementary strand to form an overhanging 5' end, using a polymerase activity, so that in both cases a double-stranded nucleic acid molecule having smooth ends results.

The term "preparing a nucleic acid molecule" encompasses any form of preparation, for example the cloning of a gene with subsequent restriction cleavage and isolation of a fragment having an overhanging or smooth end, for example, which is used as starting material for the method according to the invention. In another embodiment, the nucleic acid molecule is prepared by juxtaposition of two synthetic oligonucleotides which are at least partially complementary, wherein an overhang may result from the juxtaposition. In another embodiment single-stranded oligonucleotides are prepared.

The term "at at least one end" as used according to the invention means that synthesis may proceed unidirectionally or bidirectionally.

The "attachment" of the single-stranded nucleic acid molecules preferably occurs by hybridization. For each step of the attachment of a new single strand, the required hybridization conditions may be easily modified, if necessary, by one skilled in the art based on his technical expertise.

The single-stranded nucleic acid molecules used in the method according to the invention have a maximum length of approximately 150 nucleotides. A length between 15 and 130 nucleotides is preferred. In general, in selecting the length of the single-stranded molecules it is important to note that the yield of intact oligonucleotides in the chemical synthesis of single-stranded precursor molecules decreases with increasing length, specifically, due to faulty introduction of nucleotides. Thus, a compromise must be made between the length of the oligonucleotides and their yield. The quality of the single-stranded molecules used for the synthesis also has an influence on the yield of desired nucleic acid, using the method according to the invention. As the result of oligonucleotide purification by HPLC, the individual single-stranded nucleic acid molecules are intact for further syntheses. Lastly, the length of the oligonucleotides used in the further syntheses depends on the quantity needed for a synthesis step, and the yield from the chemical synthesis.

The term "predetermined location" as used according to the invention means that this sequence is defined either by its primary sequence or by its positioning relative to the actual cleavage site.

A predetermined location for cleavage of a nucleic acid single strand may be provided, for example, by incorporating one or more synthetic or modified nucleotides, base analogs, or a chemical group, internally or terminally, which may be cleaved using a physical, chemical, or enzymatic method, resulting in a 3'-OH end and/or a 5'-phosphate end (for example, Maxam-Gilbert reaction, etc.). Examples of nucleotides that are suitable for the method according to the invention include 5-hydroxy-2-deoxycytidine, 5-hydroxy-2-deoxyuridine, and 5-hydroxy-2'-deoxyuridine. The first two nucleotides represent substrates for *E. coli* endonuclease III and formamidopyrimidine DNA glycosylase, whereas the latter nucleotide may be cleaved using uracil DNA glycosylase and apyrimidase or alkali treatment.

For example, phosphoborane nucleotides or thioate nucleotides in the DNA sequence may bring terminal digestion of exonuclease II or T7 (gene 6) nuclease to a halt, the sequence being predetermined up to the nucleotide that is modified for the cleavage.

Another option for cleavage of the nucleic acid single strand at a predetermined location is to introduce a "mismatch" into an artificial hairpin structure. This structure may be efficiently and precisely cleaved by using a "mismatch repair" enzyme.

The particular (molecular) agent that is ultimately used as restriction activity for cleaving one or more predetermined locations in a nucleic acid double strand in the method according to the invention is not essential to the invention. Rather, for embodiments which include the cleavage of double-stranded nucleic acid it is important, as mentioned above by way of example, that the recognition sequence on the nucleic acid and the sequence that is actually cleaved are spatially separate from one another. According to the invention, the recognition sequence is generally removed by cleavage from the growing double-stranded nucleic acid molecule. The class IIS restriction endonucleases have properties that meet the requirements for such an agent. Depending on the embodiment of the method according to the

invention, representatives of this class which generate a free cohesive 3' end or an overhanging cohesive 5' end are suitable.

The properties of the restriction activities which may be used in the method according to the invention may be summarized as follows:

(a) The restricting agent may be of various types: Included are all synthetic agents which specifically cleave nucleic acids, such as synthetic peptides, peptide nucleic acid (PNA), triple helical DNA-binding oligonucleotides which are suitable for the specific processing of the nucleic acid terminus/termini within the meaning of the present invention, as well as naturally occurring DNA-cleaving enzymes. One skilled in the art is able to use (exo)nuclease or also restriction activities that are suitable for his particular purposes;
(b) These may be type IIS restriction endonucleases, for example;
(c) Asymmetrical recognition sequences (class IIS restriction endonucleases) as well as symmetrical recognition sequences may be used;
(d) As previously mentioned, the cleavage sites which are generated by the restriction activity cannot be situated within the specific recognition sequence; rather, they must be distally located at a distance from the 5' or 3' end;
(e) Removal of the interface from the recognition sequence must be precisely and unambiguously defined;
(f) To ensure the specificity of the attachment of the nucleic acid single strand in one embodiment of the method according to the invention, and to ensure an efficient ligation reaction, provided this is desired, the nuclease activity or the restricting agent preferably generates cohesive ends. This also eliminates the need, discussed above, for masking the single strands which are attached at the smooth ends, for example.

The appended reference list provides one skilled in the art with a suitable selection of restricting agents.

Carrying out step (5) or the frequency of carrying out same ultimately depends on the length of the desired end product, as well as the strand length of the available starting material.

In another particularly preferred embodiment, a synthetic, single-stranded DNA molecule (+) is generally provided whose 5' terminus is provided for hybridization with the 3' end of the single-stranded DNA molecule (–) previously synthesized based on a template. Thus, a synthetic template sequence (–) is generally provided which, starting at the 3' end (+) of the previously synthesized strand, is available for the synthesis of a new DNA strand (+). This results in a double-stranded 3' terminus which terminally masks the double strand. To simplify the subsequent hybridization reaction, the complementary template strand may be completely (Figure 8A) or partially (Figure 8B) degraded, using a 5'-3' exonuclease activity, for example. Internal masking of the single-stranded molecule (using borane phosphates, for example) is then used to precisely define the stop point of the exonuclease (K.W. Porter

et al., *Nucl. Acids Res.* 25(8), 1611-1617 (1997)).

Advantages offered by the present invention compared to the prior art include the following, among others:

1. Availability:

Sequences of nucleic acids—genes, for example—are widely available and may be obtained quickly (within days), provided that the nucleotide sequences are known from databases, for example. With knowledge of these sequences, combined with the teaching according to the invention, one skilled in the art may synthesize any desired nucleotide molecule.

2. Storage and transport:

Nucleic acid molecules in the size range of genes no longer have to be physically stored in refrigerators, which have high energy consumption. Their sequences may be managed in data processing units, and if necessary may be easily synthesized in a biological system. This also eliminates the need for transport, since DNA sequences may be sent via e-mail.

3. Capability for manipulation:

Any given nucleotide sequences such as DNA sequences and gene sequences, for example, are mutated in vitro in order to achieve certain properties, such as heat stability, changes in pH, or solubility in certain solvents, as well as to optimize or alter the biological behavior of their gene products. In vitro mutagenesis is a complicated task despite a variety of different methods. As a result of the synthesis options, it is also possible to insert any desired number of mutations, even at widely separated sequence positions, since the sequence may be freely defined and determined in advance. Thus, any desired number of variants may be produced.

The capability for freely synthesizing the DNA sequences will transfer many of the currently used methods of time-consuming DNA manipulations from the laboratory to the synthesis machine, resulting in great cost and time savings. The experiment to be conducted by one skilled in the art then involves designing a nucleic acid sequence on a computer editor and checking whether the desired properties due to the sequence manipulations result in the biological model or in vitro. The method according to the invention therefore contributes to the further development of reverse genetics techniques.

In one preferred embodiment of the method according to the invention, additional nucleic acid molecules, fragments thereof, and/or nucleotides which are not incorporated into the prepared nucleic acid molecule are separated after step (2), (2a), (3), (4), (4a), (5), and/or (5a).

Although the separation of the unincorporated single-stranded nucleic acid molecules is preferred, it is not absolutely necessary, and may be achieved by one skilled in the art using standard processes such as column chromatography methods. The concentration of free nucleotide phosphates could possibly be limiting, in particular for the overall yield of desired nucleic acid, and consumed nucleotides could

interfere with the synthesis and ligation reaction which is necessary for carrying out the method according to the invention, for example for the generation of smooth ends, or the attachment of single strands of nucleic acid at smooth ends and subsequent generation of the complementary strand, as described above. A high concentration of various single-stranded DNAs increases the risk of undesired by-products. As a practical matter, it is therefore advantageous when each individual synthesis step is able to proceed under optimal conditions. It is therefore recommended that in each case the unneeded single strands be separated before the next synthesis step, for example in a matrix-coupled reaction, after which consumed nucleotides and excess single-stranded nucleic acids are eluted. Of course, the separation may be carried out during or after step (5).

The above-described optional ligation in the embodiment of the method according to the invention, which includes the attachment of complementary end-position nucleotides by hybridization, may be carried out, for example, before, concurrently with, or after step (4). In another embodiment it may be carried out during or after step (5). In one example of the ligation after step (5), bacteria, *E. coli*, for example, are transformed using the unligated synthesis product, and the ligation is performed by endogenic ligases. It is known that with increasing size of the complementary overlapping cohesive ends it is possible to transform *E. coli*, for example, using suitable DNA and making use of the endogenic ligase activity for circularization. In this regard, gaps and overhanging single-stranded DNAs are in fact tolerated, since repair mechanisms restore the integrity of the circular double strand. Single-strand areas are replenished and repaired when at least one phosphodiester backbone is intact. A ligation is preferably conducted when the overhangs have a length of only a few nucleotides. For longer overhangs, gaps may appear between the end-position nucleotides of the single strand and the double strand, which are closed using a polymerase activity, for example, before a ligation reaction. Since the restriction enzymes known thus far usually generate only relatively short cohesive ends, a C and G or A and T "tailing" with terminal transferase is also conceivable, which produces long overlap areas that may be directly transformed without in vitro ligation. Lastly, the synthesis product may also be provided for ligation after the step of isolating the completed synthesis product.

As previously mentioned, in one preferred embodiment of the method according to the invention the predetermined location of the nucleic acid molecule may be provided in an artificial hairpin structure by incorporating a synthetic or modified nucleotide, a base analog, a chemical group, or a "mismatch" which may be cleaved using a physical, chemical, or enzymatic method.

In one particularly preferred embodiment, the synthetic or modified nucleotide is 5-hydroxy-2-deoxycytidine, 5-hydroxy-2-deoxyuridine, or 5-hydroxy-2'-deoxyuridine.

As also previously mentioned, in another preferred embodiment the present invention relates to a method in which the linkage of two end-position nucleotides is achieved via a 3'-hydroxy end and a 5'-phosphate end, using a ligase activity, and the attachment is achieved by the hybridization of complementary sequences.

In another preferred embodiment of the method according to the invention the nucleic acid is DNA. In another preferred embodiment of the method according to the invention the nucleic acid is RNA. The generation of DNA/RNA hybrids is also encompassed by the method according to the invention.

In another preferred embodiment of the method according to the invention, the masking in step (3) is carried out additively and subtractively by adding or removing a chemical group or a chemical molecule. In one preferred embodiment of the method according to the invention, the masking of a 5' end is carried out by removing the phosphate group(s) or by introducing a 5'-modified nucleotide (for example, biotin-dNTP, digoxygenin-dNTP, etc.). As mentioned above, masking the 5' end of a single-stranded nucleic acid molecule to be attached in the corresponding embodiment of the method according to the invention prevents an undesired ligase side reaction between the 5' and 3' ends of the single-stranded nucleic acid molecules, and therefore prevents the possible formation of concatemers, thus ensuring an optimal yield from the method following the teaching according to the invention.

In one particularly preferred embodiment of the method according to the invention, the masking is carried out by introducing at least one 5'-modified nucleotide.

In another preferred embodiment of the method according to the invention, as previously mentioned, a masked 3' end is characterized by the presence of an amino block, a dideoxynucleotide, a 3'-phosphate, or a synthetic 5' end.

In another preferred embodiment, the additional nucleic acid molecule forms a hairpin loop, at the prepared nucleic acid molecule after attachment and/or linkage removed end, which is used as a primer for the polymerase activity.

In another preferred embodiment, the invention relates to a method in which the cleavage takes place in step (4) at a predetermined location, using a triple helical DNA which cleaves in a sequence-specific manner.

A triple helical DNA is formed, for example, when a single-stranded DNA having a heavy metal (HM) coupled at the end is attached to a DNA double strand and, if the sequence conditions are suitable, a triple helical structure having a DNA double strand is formed. The nucleic acid double strand is cleaved by the heavy metal at a defined position.

Furthermore, as previously mentioned, according to the invention any particular physical, chemical, or enzymatic nucleic acid cleavage may be used which is conducive to the attachment of a single-stranded nucleic acid molecule for the subsequent ligation with the double-stranded nucleic acid molecule. Further examples of such include methodological approaches based on designed peptides or peptide nucleic acid (PNA). One skilled in the art may obtain an overview of the above-mentioned molecules and

examples of their possible uses from the appended reference list.

Another preferred embodiment of the invention relates to a method in which the cleavage takes place in step (4) at a predetermined location, using a type IIS restriction endonuclease. Type or class IIS enzymes have an asymmetrical, i.e., nonpalindromic, recognition sequence. The cleavage sites are located at either the distal 5' or 3' end with respect to the recognition sequence. Either 5' (BspMI, for example) or 3' (RleAI, for example) overhanging ends or smooth ends (BsmFI, for example) are generated.

In one particularly preferred embodiment of the method according to the invention, the type IIS restriction endonuclease is the RleAI enzyme from *Rhizobium leguminosarum* (Vesely Z., Müller A., Schmitz G., Kaluza K., Jarsch M., Kessler C. (1990) R1eAI: A novel class IIS restriction endonuclease from *Rhizobium leguminosarum* recognizing 5'-CCCACA(N12/9) – 3', gene 95: 129-131).

In another preferred embodiment of the method according to the invention, the double-stranded nucleic acid molecules and/or the single-stranded nucleic acid molecules are of synthetic or semisynthetic origin; the use of synthetic single-stranded molecules is particularly preferred for the synthesis.

Semisynthetic molecules may be prepared by introducing nucleic acid fragments, obtained from DNA (dsDNA, ssDNA) or RNA amplified in vivo (bacteria, yeast) in one or more intermediate steps of the synthesis according to the invention, at defined locations by means of ligation reactions. In some cases this strategy may contribute to considerable cost reductions. For example, the nucleic acid molecule [used] as starter molecule may likewise be a DNA molecule prepared in vivo, to which any desired DNA sequences are attached by recursive DNA synthesis.

In another preferred embodiment of the method according to the invention, the synthesis is at least partially automated. Thus, for example, in a nucleic acid (gene) automatic synthesizer for synthesizing nucleic acid double strands from nucleic acid single strands a battery of automated chemical oligonucleotide syntheses (a technology already in widespread use) is able to provide the raw material for synthesizing biologically active double-stranded DNA molecules (entire genes, for example), which are produced from the chemically synthesized oligonucleotides in a likewise automated method.

The double-stranded nucleic acids to be extended are bound to the synthesis matrix in a synthesis chamber. The same steps as described above proceed in a cyclical reaction sequence in this synthesis chamber. The reaction by-products of the preceding reaction are flushed from the synthesis chamber before starting a new reaction. The starter molecule which is extended by one nucleic acid molecule remains bound to the synthesis matrix. In each synthesis step a nucleic acid having a different sequence is introduced, ultimately resulting in an optionally double-stranded nucleic acid having the desired nucleotide sequence.

In one particularly preferred embodiment, the invention relates to a method in which the synthesis is carried out bound to a matrix.

All substrate materials to which a nucleic acid is able to bind, and whose properties are compatible with the desired recursive nucleic acid synthesis, are suitable as a synthesis matrix, for example streptavidin-coated surfaces, wherein the double-stranded nucleic acid molecule used as starter molecule is coupled to the synthesis matrix via an introduced biotinylated nucleotide. Other preferred synthesis matrices include Nylon surfaces to which sequences containing poly dT are coupled by UV irradiation, as well as tosyl-, active ester-, or epoxy-activated surfaces (GOPS, for example), the binding preferably taking place via an "amino link" such as glass (CPG, glass wool, etc.), silicate, latex, polystyrene, epoxy, or silicon.

In another preferred embodiment of the method according to the invention, the synthesized nucleic acid molecule is isolated after the synthesis.

This is carried out on the one hand by introducing an affinity-enhancing agent, for example biotin, digoxiginin, a histidine tag, or a maltose radical, in the last synthesis step. The synthesis end products labeled in this manner may thus be easily and cost-effectively isolated using appropriate columns.

Alternatively, in the last synthesis step of the method according to the invention a plasmid is linked to the synthesis product, and the resulting nucleic acid molecule, optionally after being recircularized in bacteria, is introduced and duplicated. Alternatively, defined sequences to which primers are able to specifically bind are incorporated into the prepared nucleic acid molecule and the nucleic acid molecule used in the last synthesis step. With the aid of these primers, the completely synthesized nucleic acid molecule may be amplified in a PCR reaction, thus isolating the synthesis product from the affinity matrix. If sequence motifs (type IIS recognition sites, for example) are present in the terminal sequences by means of which cohesive ends may be generated, even higher-order sequence motifs may be easily synthesized from individual DNA molecules.

In another preferred embodiment of the method according to the invention, single-stranded nucleic acid molecules are isolated by denaturing the double-stranded nucleic acid molecule.

This embodiment of the method according to the invention is suitable for preparing single-stranded nucleic acid molecules of any desired composition. In this regard, special mention is made of the possibility for preparing such RNA molecules.

Lastly, the invention relates to a kit comprising:

(a) a ligase, and/or
(b) a polymerase,
(c) optionally a type IIS restriction enzyme,

(d) optionally a uracil DNA gycosylase and an apyrimidase, and/or an endonuclease III and a formamidopyrimidine DNA glycosylase, and/or a "mismatch repair" enzyme,
(e) optionally a phosphatase, a terminal transferase, and/or an exonuclease,
(f) optionally a wash buffer for eluting reaction by-products and material not introduced into the product of the synthesis according to the invention,
(g) optionally a synthesis matrix, using a nucleic acid molecule which is optionally already bound thereto as starter molecule,
(h) optionally suitable reaction buffers for the enzymes listed in (a) through (e).

Based on the teaching of the present invention as well as the general technical knowledge in this field, the manufacturer of the kit according to the invention is aware of how to produce and formulate the individual components of the kit, for example the buffers. The kit according to the invention may also optionally contain a starter molecule which is not bound to a matrix and/or a set of suitable single-stranded molecules.

Description of the figures

Figure 1. In vitro ssDNA synthesis in the 3' to 5' direction. (1) A starter molecule (n) coupled to a matrix is linked about an n+1th single-stranded molecule by a 3'-5' phosphodiester bond, using a ligase activity. The n+1th ssDNA has a terminal uracildeoxy nucleotide. [(2)] The glycosidic bond of the base uracil is cleaved by the DNA uracil glycosylase, resulting in an apyrimidinic position. This in turn is cleaved by an apyrimidinic endonuclease activity (exonuclease III), resulting in a 5'-phosphate end and a 3'-OH end. (3) The n+2th ssDNA molecule is linked to the released 5'-phosphate end in the n+2th ligation reaction. A subsequent phosphatase reaction (not shown; see Figure 4) inactivates all DNA chains for the n+3th ligation step [and] for all subsequent ligation steps, provided that no n+2th ssDNA molecule has been introduced in the n+2th step. As a result of the DNA uracil glycosylase and the apyrimidinic endonuclease activity, a 5'-phosphate may once again be provided for the next reaction sequence (n+3) by processing. All steps are repeated k times until the last ssDNA molecule has been introduced in the mth step.

Figure 2. In vitro dsDNA synthesis in the 3' to 5' direction. All steps are carried out as in Figure 1, except that in the last step a polymerization reaction is initiated, starting at the 3' end of the starter molecule, which replenishes the newly synthesized single strand to form a double strand. Alternatively, in both the first and last step a primer of a primer pair may be introduced, thus producing a double-stranded molecule by amplification, using the PCR reaction.

Figure 3. In vitro dsDNA synthesis in the 3' to 5' direction. All steps are carried out as in Figure 1. Within each synthesis cycle, after the phosphatase treatment a polymerization step is initiated which converts the ligated ssDNA to dsDNA. The processing may also be carried out using a type IIS restriction endonuclease, provided that a recognition sequence has been introduced into each of the ssDNA

fragments.

Figure 4. In vitro dsDNA synthesis in the 3' to 5' direction. All steps are carried out as in Figure 1. A phosphatase reaction carried out after a ligation step inactivates all DNA molecules for the next ligation step and for all subsequent ligation steps, provided that no ssDNA molecule has been introduced in the n–1th synthesis cycle. As a result of the DNA uracil glycosylase and the apyrimidinic endonuclease activity, in each synthesis cycle a 5'-phosphate may once again be provided for the next reaction sequence by processing. All steps are repeated k times until the last ssDNA molecule has been introduced in the mth step.

Figure 5. In vitro ssDNA synthesis in the 3' to 5' direction. (1) A starter molecule (n) is linked about an n+1th single-stranded molecule by a 3'-5' phosphodiester bond, using a ligase activity. The n+1th ssDNA has a terminal uracildeoxy nucleotide, and is 5'-phosphorylated and 3'-blocked (-X). [(2)] The glycosidic bond of the base uracil is cleaved by the DNA uracil glycosylase, resulting in an apyrimidinic position. This in turn is cleaved by an apyrimidinic endonuclease activity (exonuclease III), resulting in a 5'-phosphate end and a 3'-OH end. (3) The n+2th ssDNA molecule is linked to the released 5'-phosphate end in the n+2th ligation reaction. A subsequent terminal transferase reaction with a dideoxy trinucleotide (not shown; see Figure 7) inactivates all DNA chains for the n+3th ligation step [and] for all subsequent ligation steps, provided that no n+2th ssDNA molecule has been introduced in the n+2th step. As a result of the DNA uracil glycosylase and the apyrimidinic endonuclease activity, a 3'-OH may once again be provided for the next reaction sequence (n+3) by processing. All steps are repeated k times until the last ssDNA molecule has been introduced in the mth step.

Figure 6. In vitro ssDNA synthesis in the 5' to 3' direction. (1) A starter molecule (n) is linked about a n+1th single-stranded molecule by a 3'-5' phosphodiester bond, using a ligase activity. All other steps are carried out as illustrated in Figure 5.
In the last step, initiated by a 3'-terminal hairpin structure, for example, a 3' end for a DNA polymerization reaction may be provided, or a dsDNA polymerization may be carried out as described in Figure 2.

Figure 7. In vitro ssDNA synthesis in the 5' to 3' direction. (1) Illustration of the reaction for inactivation of unligated ends. A terminal transferase reaction with a dideoxy trinucleotide inactivates all DNA chains for the next ligation step [and] for all subsequent ligation steps, provided that no ssDNA molecule has been introduced in the n–1th synthesis step. As a result of the DNA uracil glycosylase and the apyrimidinic endonuclease activity, a 3'-OH may once again be provided for the next reaction sequence by processing. All steps are repeated k times until the last ssDNA molecule has been introduced in the mth step

Figure 8. Polymerase-based DNA synthesis. A starter molecule (n) is provided. Synthetic oligonucleotides are sequentially provided in a cyclical reaction sequence, the 5' end thereof being

provided for hybridization with the respective preceding 3' end of the complementary DNA strand. The synthesis to form the double strand starts from the 3' end. Figure 8A shows a complete degradation of the template molecule using T7 (gene 6), whereas Figure 8B shows a partial degradation using exonuclease III.

The present invention is explained using the following examples.

Example 1

The in vitro recursive DNA synthesis may be used for manipulating DNA sequences in vitro. On the one hand, gene mutations such as deletion mutagenesis as well as multiple simultaneous deletions in a gene, gene fusions with creation of new properties, insertion mutagenesis, substitution mutagenesis, and sequence inversions may be carried out. In addition, one to any desired number of point mutations may be introduced into a sequence. All DNA sequences may be directly prepared in parallel syntheses without intermediate cloning steps.

The functional alterations of the in vivo biological activity resulting from the sequence manipulations on the one hand may act at the protein level, provided that the encoding sequences may be translated in functional proteins. The method may thus be used for implementing concepts for enzyme or protein design.

On the other hand, it is possible to manipulate the DNA sequences of regulatory *cis* elements in order to alter the binding activity of transactivators and suppressors, to investigate the behavior thereof, or even to create entirely new combinations of *cis* elements. In addition, the activity of RNA molecules (ribozymes, for example) could also be manipulated, provided that the manipulated DNA is transcribed.

The following example for use of the in vitro recursive DNA synthesis method involves the manipulation of DNA sequences in order to analyze the binding activity of a transactive regulator protein at a bacterial promoter region. The in vitro mutagenesis of the binding sites is used to investigate the effects on the DNA-binding protein. The DNA wild type sequence and the mutants of the *cis*-active element are illustrated in Figure 7, and the sequence manipulations are explained in the text. The object of the tests was to allow investigation of the functionality of the binding of the regulator in a different sequence context in vitro, and possibly also in vivo.

Palindromic sequence segments are present on a Sau IIiA DNA fragment in the region of a bacterial promoter, the structure of which is very similar to sequences which also take part in transcriptional regulation in other systems. The transcriptional activity of the S^{70}-like promoter located in the 5' sequence region of the 6-HDNO gene, on which the *cis*-active elements reside, has been studied in detail by Mauch et al. (1990) in the in vivo heterologous *E. coli* system. The cloning of this DNA fragment, which was used in most of the DNA binding studies of the referenced work, is referred to

below as "WT-6-HDNO promoter fragment" (Figure 7A (1–3)).

The incubation of crude extracts from *Arthrobacter nicotinovorans* cells (10–40 mg total protein) having a radioactively labeled 6-HDNO binding fragment of the 6-HDNO gene in the presence of a competitor (generally poly(dI-dC)) from the promoter region showed, after separation of the constituents of this incubation assay in the electrical field of a native PAA gel, clear retention of the DNA fragment compared to a control assay without addition of crude extract protein.

The experimental approach used to identify the NicR1 binding activity is referred to as gel retention analysis. This method may be used for in vitro qualitative and quantitative investigation of the kinetic and functional behavior of DNA binding proteins as a function of various parameters. In addition, under certain conditions conclusions may be drawn concerning the structure of the DNA/protein complex.

Using crude extracts, in the gel retention experiment it is generally possible to identify a dominant retarded band. In some cases a second band may also be identified. This DNA binding activity was suppressed not by large quantities of unlabeled nonspecific competitor DNA, but, rather, was more likely suppressed by unlabeled binding fragment in very small quantities. It was therefore assumed that this DNA binding activity is specific, and is correlated with the transcriptional regulation of the nicotine regulon. This regulon is abbreviated as NicR1 (nicotine regulator 1) (Mauch et al., Bernauer et al., 1992).

The behavior of the nicR1 binding activity in the gel retention experiment in the referenced study was analyzed in order to draw conclusions concerning the location, specificity, kinetics, and stoichiometry of the binding reaction, and to investigate the response of the DNA binding function to manipulations at the WT-6-HDNO promoter fragment, and to potential effector substances.

The objective of these tests was to provide information concerning which molecular mechanisms could be responsible for the regulation of the 6-HDNO gene. In addition, the culturing conditions and the induction status of the *Arthrobacter nicotinovorans* cells were varied, ultimately resulting in the preparation of crude extract for analysis in the gel retention experiment. These experiments were designed to clarify whether the binding behavior of NicR1 changes, or whether additional factors could be detected as a function of one of the test parameters. The standard reaction buffer used for the protein/DNA binding tests was based on the reaction buffer used by Garner and Revzin (1981). The NicR1 binding activity may be augmented by ammonium sulfate fractionation. The augmentation of the NicR1 binding activity was the prerequisite for tests to analyze the binding behavior of NicR1 with simultaneous binding of both palindromic sequences which are present on the WT-6-HDNO promoter fragment.

The WT-6-HDNO promoter fragment from the 5' control region of the 6-HDNO gene has very interesting sequence characteristics (see Figure 7). Extended inverted sequence repetitions (IR) and other

noteworthy sequence motifs are present. Characteristic sequence arrangements within the 6-HDNO gene promoter region are shown in Figure 7. Shown are two inverted repetitions, IR1 and IR2, having extensive homology to one another (Figure 7). The right palindromic half side of IR2 is repeated once again in the 5' region. Such palindromes are structural features found in many bacterial *cis*-active regulator regions.

IR1 and IR2 are separated by an interpalindromic sequence that is 50 bp in length. The palindromic half sides of IR1 are homologous to one another over 17 bp, and for IR2, over 9 bp. The palindrome of IR1 has an extension which is larger by 12 base pairs, but in this region shows two insertions of two base pairs and one base pair (AT, A) each. Ten to 12 base pairs of IR1 in the 5' half, and 9 of 12 base pairs in the 3' half of the sequence, are homologous to IR2 (Figure 7A, sequences of IR1 and IR2). These sequence-specific features may have structural and functional significance for the binding of the "in trans"-binding active protein NicR1 and an S^{70}-like RNA polymerase. IR1 represents an almost perfect S^{70}-like promoter sequence having a noteworthy modification. The –10– region differs from the consensus sequence TATAAT by the insertion of a C, resulting in the sequence TAT-<u>C</u>AAT. A –30– region is present in the sequence of IR2, but there is no similarity to the known –10– region of an S^{70}-like promoter sequence. The distance of 16 bp between the –10– region and the –30– region corresponds to the S^{70} ideal distance of 17 [bp]. The –35– region of an S^{70}-like promoter is integrated into the sequence of IR2, and a consensus-like –10– region is absent. Several other sequence characteristics could also further reflect the activity of transactive regulatory elements at the 5' sequence region of the 6-HDNO gene. Three NlaI-(CATG) recognition palindromes are present at a homologous position within the palindromes IR1 and IR2. It is interesting that such an interface is also present behind the left palindromic half side of IR2, at the nonhomologous position. This raises the question of coincidence or necessity of such a structure. Likewise noteworthy sequence motifs are present outside the palindromic sequences of IR1 and IR2. GC- and AT-rich sequences are arranged in alternation. An interesting sequence characteristic of these domains is the presence of GC-rich sequence segments which are interrupted by an AT-rich sequence segment in the 6-HDNO-5' sequence. The GC sequence blocks are located above the 5' region of the S^{70}-like promoter. A detailed base neighborhood analysis according to the algorithm described in Ebbole and Zalkin (1989) shows that this sequence for the most part is not statistically significant. To demonstrate this, the present applicant wrote a Pascal computer program. The sequences within the palindromic regions are composed of apparently regularly alternating short GC and AT regions having a very balanced GC content, whereas larger sequence segments having a very unbalanced GC content are present in the 5' direction with respect to the palindromes (Figure 7A). The GC content initially increases outside palindrome IR2 from the direction of 5', first passing through a GC maximum, and then a GC minimum (Figure 7A). The situation is repeated in front of palindrome IR1. Alternating GC- and AT-rich sequence segments are associated with structural properties of protein binding. The AT-rich positions are turned inwardly toward the protein at their small DNA groove, whereas the GC-rich sequence blocks point outwardly. GC-rich promoters, which no longer have any sequence similarity to the known S^{70}-like promoters, are present in Streptomyces species.

As starter molecule (see Figure 7AA (0)) for the recursive DNA synthesis, the plasmid pUC19 (Yanish-Perron et al., 1985) was subjected to double digestion using BamHI and KpnI, and was purified through an agarose gel. KpnI has the recognition sequence 5'-GGTAC'C-3'. An oligonucleotide having a complementary sequence was attached to the 3' overhanging KpnI end in the presence of a T4 ligase, T4 DNA polymerase, and 0.2 mM dNTP at standard conditions (Sambrook et al., (1989)), ligated, and replenished to form a double strand. The oligonucleotide has the recognition sequence of the restriction endonuclease RleAI plus several additional bases at the 5' end (see Figure 7A (1)). The DNA molecule which is now double-stranded (replenished overhanging synthetic oligonucleotide) was restricted with an enrichment fraction of the restriction endonuclease RleAI from *Rhizobium leguminosarum* (Figure 7 (2) and (3'), respectively). The reaction conditions were taken from Veseley et al. (1990). This enzyme produces 3' overhanging ends outside its asymmetrical binding site. This specificity is thus far unique, and allows the repeated attachment of an oligonucleotide and the priming for a DNA polymerization. The short DNA fragment having the RleAI recognition sequence was separated from the plasmid by purification through an agarose gel. At the overhanging 3' end resulting from the restriction reaction, a oligonucleotide complementary to this end was reattached (Figure 7A (2)) and replenished, as mentioned above (also see Figures 1 and 2). The same reaction was carried out with the oligonucleotide (Figure 7A (3)) and the variants (Figures 7B-1 through B-7). By using synthetic oligonucleotides it was possible to produce in parallel seven sequence variants and the wild type sequence. Following the reaction sequence (Figure 7A (3)), the newly produced DNAs were cleaved with BamHI (see sequence, Figure 7A(3")), the vector (pUC19 + binding fragment) was circularized, and transformation into *E. coli* was carried out according to standard methods. If the RleAI recognition sequence has a terminal location at the end also for the synthetic oligonucleotides, the DNA synthesis reaction as in this example may be extended by one step in each case.

To characterize the binding properties of NicR1, sequence alterations were introduced into the IR1 binding site bearing the S^{70}-like promoter, and the length of the interpalindromic sequence (IS, Figure 7B-5, -6, -7) was varied. The objective of the latter tests was to investigate the steric requirements for the palindromic binding sequence IR1. Sequence modifications which were introduced into the WT-6-HDNO promoter fragment are shown in Figure 7.

The alterations introduced by in vitro recursive DNA synthesis into the sequence of the promoter-containing IR1 palindrome and the interpalindromic sequence are illustrated in Figure 7B. The reduction of IR1 to an octamer (Figure 7B-3) as well as deletion of the central G position (Figure 7B-4) destroy the capability of NicR1 to bind to IR1.

When the cleavage products are used in the gel retention experiment, only IR2 is retarded, but not the mutated fragment containing IR1. The construct illustrated in Figure 7B-4 shows retention only via the binding of NicR1 to IR2. Since the size of the complex at IR2 is the same as the complex at IR1, this indicates that the same protein is bound to both palindromes.

Contrary to the pronounced effect on the NicR1 binding resulting from the changes to the length and to the symmetry of the palindrome IR1, changes in the number of helical windings in the interpalindromic sequence resulted in no difference in the NicR1 binding at both palindromes. The length of the interpalindromic sequence was altered by deletions as well as insertions of 5 bp in each case (Figure 7B-6 and -7). These alterations each correspond to one-half of a helical winding. As a result, in these DNA mutants the IR2 binding site is twisted by 180° relative to the IR1 binding site. In addition, the interpalindromic sequence 50 bp in length has been decreased by 20 bp (Figure 7B-5). The pattern of the gel retention experiment reflecting these changes (Figure 7B-5, -6, -7) was identical to the control pattern observed for the unaltered 6-HDNO promoter fragment 242 bp in length (Figure 7B-1).

The right half of IR1 contains the –10– region of the promoter of the 6-HDNO gene, which differs from the consensus sequence of the promoter of the S^{70} RNA polymerases as the result of insertion of an additional cytosine-containing base position in the TATAAT sequence (Figure 7B-1). The question arises as to whether this unusual S^{70} –10– region contributes to the specificity of the NicR1 binding to IR1. In the gel retention experiment (Figure 7B-2) using NicR1, deletion of the cytosine radical at the corresponding position (Figure 7B-2) resulted in no change in the protein binding pattern compared to the pattern obtained from the unaltered DNA fragment, most likely for the binding of the S^{70}-like RNA polymerase of *E. coil*.

The conclusion that the two mutations in Figure 7B-3 and -4 greatly reduce, if not totally eliminate, the capability of NicR1 to bind to the palindrome IR1 is supported by further tests not described here.

Example 2

Example for the synthesis of:

PLASMIDS π-AN7 885 bp : Huang, Little, Seed (1985) in : Vectors: A survey of molecular cloning and their applications", Rodriguez, R., ed., Butterworth Publishers, Stoneham, MA, USA.

STARTER MOLECULE:

AAUGCGGCCGCTCACGAGCCGCGCGGTTAATTAACTCGAGAABTCCGCGGTGCAATTAATT-X

Restriction enzymes : : EagI, Bst2BI, AccBSI, NotI, PacI, XhoI, EcoRI, SacI

B=Biotin
X=AminoBlock

π-AN7- SEQUENCE : Source : GeneBank

POSITION 3 : Uracil

O1 AAUTTTCGGACTTTTGAAAGTGATGGTGGTGGGGGAAGGATTCGAACCTTCGAAGTCGATGAC3'

O2 AAUGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCACGGGTAATGCTTTT3'

O3 AAUACTGGCCTGCTCCCTTATCGGGAAGCGGGGCGCATCATATCAAATGACGCGCCGCTGTAA3'

O4 AAUAGTGTTACGTTGAGAAAGAATTCCCGGGGATCCGTCGACCTGCAGATCTCTAGAAGCTTG-3'

O5 AAUCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT-3'

O6 AAUCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA-3'

O7 AAUGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC-3'

O8 AAUTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT-3'

O9 AAUAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCG-3'

O10 AAUCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG-3'

O11 AAUCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT-3'

O12 AAUTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGC-3'

O13 AAUTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG-3'

O14 AAUCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTC-3'

O15 AAUAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAAATTC-3'

LAST SYNTHESIS CYCLE US:

B=Biotin-AAUCTCGAGAATTCCGCGGTGCAATTAATTAAAAAAAAAAAAA

Underscore : SupF
BlueSequence :Polylinker
BlackSequence :ColEI

Brief protocol:

A µg-biotinylated starter molecule was bound to streptavide-coated Dynabeads. 0.2 mM biotin deoxyuracil was then adjusted (0.5 h incubation at RT) in order to block all biotin binding sites.

In 16 synthesis cycles (ligation, T4 RNA ligase; uracil DNA glycosylase, exonuclease III; phosphatase) the 17 DNA molecules were linked to form a single-stranded DNA which included the entire plasmid sequence.

The ssDNA was replenished to form dsDNA, using T4 DNA polymerase, starting from the 3' end of the starter molecule. NotI was used to release the dsDNA from its binding to the Dynabeads.

The same quantity of fresh Dynabeads was added.

Only the molecules containing biotin were bound to the column. Molecules not containing biotin from the last ligation reaction were flushed away.

Cleavage was carried out using the restriction endonuclease PacI.

The Dynabeads were pelletized using a magnet.

The molecules were precipitated from the supernatant using ethanol, and the molecules containing the T4 DNA ligase were circularized.

The circularized synthetic plasmid molecules were then transformed into *E. coli* DH10α/P3 according to standard protocol, and were selected against Tet/Amp on LB plates.

Reference List

Bernauer, H., Mauch, L. and Brandisch, R. (1992) Interaction of the regularory protein NicR1 with the promotor region of the pAO1-encoded 6-hydroxy-D-nicotine oxidase gene of Arthrobacter oxidans, Mol Microbiol 6: 1809-1820.

Bigey, P., G. Pratviel, et al. (1995). "Cleavage of double-stranded DNA by 'metalloporphyrin-linker-oligonucleotide' molecules: influence of the linker. Nucleic Acids Res 23(19): 3894-900.

Francois, J. C. and C. Helene (1995). "Recognition and cleavage of single-stranded DNA containing hairpin structures by oligonucleotides forming both Watson-Crick and Hoogsteen hydrogen bonds. Biochemistry 34(1): 65-72.

Gao, X., A. Stassinopoulos, et al. (1995). "Structural basis for the sequence-specific DNA strand cleavage by the enediyne neocarzinostatin chromophore. Structure of the post-activated chromophore-DNA complex. Biochemistry 34(1): 40-9.

Garner, M.M., and Revzin, A. (1981) A gel elektrophoresis method for quantifiying the binding of proteins to specific DNA regions: application to components of the E. coli lactose operon regulatory system. Nucleic Acids Res. 9:3047-3060.

Harford, C., S. Narindrasorasak, et al. (1996). "The designed protein M(II)-Gly-Lys-His-Fos(138-211) specifically cleaves the AP-1 binding site containing DNA. Biochemistry 35(14): 4271-8.

Kane, S. A., S. M. Hecht, et al. (1995). "Specific cleavage of a DNA triple helix by FeII-bleomycin. Biochemistry 34(51): 16715-24.

Mauch L., Bichler V., and Brandisch R (1990) Funktional analysis of the 5' regulatory region and the UUG translation initiation codon of the Arthrobacter oxidans 6-hydroxynicotine oxidase gene. Mol Gen Genet 221: 427-434.

Norton, J. C., J. H. Waggenspack, et al. (1995). "Targeting peptide nucleic acid-protein conjugates to structural features within duplex DNA. Bioorg Med Chem 3(4): 437-45.

Oikawa, S., M. Kurasaki, et al. (1995). "Oxidative and nonoxidative mechanisms of site-specific DNA cleavage induced by copper-containing metallothioneins. Biochemistry 34(27): 8763-70.

Sambrook, J. Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press 1:1,25-1.28.

Sergeev, D. S., T. S. Godovikova, et al. (1995). "Cleavage of a double-stranded DNA target by bleomycin derivatives of oligonucleotides, forming a ternary complex. Bioorg Khim 21(3): 188-96.

Vesely Z., Müller A, Schmitz G, Kaluza K, Jarsch M, Kessler C (1990) *RleAI* : a novel class - IIS restriction endonuclease from *Rhizobium leguminosarum* recognizing 5'-CC-CACA(N12/9)-3', Gene 95: 129-131.

Yanish-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotid sequence of the M13mp18 and pUC19 vectors, Gene 33: 103-109.

Claims

1. Method for synthesizing nucleic acid molecules, including the following steps:
   (1) Preparing a nucleic acid molecule that has at least one end which allows an attachment and/or linkage to an additional nucleic acid molecule;
   (2) Attaching and/or linking at least one additional nucleic acid molecule to the nucleic acid molecule, wherein one end of the at least one additional nucleic acid molecule is attached and/or linked to the at least one end of the nucleic acid molecule, and for the case of a linkage, the other end of the at least one additional nucleic acid molecule is masked;
   (3) Masking the at least one end of the nucleic acid molecule which has not been attached and/or linked to an additional nucleic acid molecule;
   (4) Cleaving the at least one additional attached and/or linked nucleic acid molecule at a predetermined location, wherein the masking is removed, and one end is generated which allows an attachment and/or linkage to an additional nucleic acid molecule; and
   (5) Repeating steps (2) through (4) at least once, optionally multiple times, wherein in each case suitable nucleic acid molecules are used in step (2).

2. Method according to Claim 1, wherein the additional nucleic acid molecule is a single-stranded molecule.

3. Method according to Claim 2, which includes the following step after step (2):
   (2a) Replenishing the second nucleic acid strand, which has a sequence that is complementary to the single strand, using a polymerase activity, the masking optionally being removed beforehand.

4. Method according to Claim 2, which includes the following step after step (4) or (5):
   (4/5a) Replenishing the second nucleic acid strand, which has a sequence that is complementary to the single strand, using a polymerase activity.

5. Method according to one of Claims 1 through 4, wherein additional nucleic acid molecules, fragments thereof, and/or nucleotides which are not incorporated into the prepared nucleic acid molecule are separated after step (2), (2a), (3), (4), (4a), (5), and/or (5a).

6. Method according to one of Claims 1 through 5, wherein the predetermined location of the nucleic acid molecule is provided [in] an artificial hairpin structure by incorporating a synthetic or modified nucleotide or a base analog which may be cleaved using a physical, chemical, or enzymatic method.

7. Method according to Claim 6, wherein the synthetic or modified nucleotide is 5-hydroxy-2-deoxycytidine, 5-hydroxy-2-deoxyuridine, or 5-hydroxy-2'-deoxyuridine.

8. Method according to one of Claims 1 through 7, wherein the linkage of two end-position nucleotides is achieved via a 3'-hydroxy end and a 5'-phosphate end, using a ligase activity, and the attachment is achieved by the hybridization of complementary sequences.

9. Method according to one of Claims 1 through 8, wherein the nucleic acid is DNA or RNA.

10. Method according to one of Claims 1 through 9, wherein the masking in step (3) is carried out additively or subtractively by adding or removing a chemical group or a chemical molecule.

11. Method according to Claim 10, wherein the masking is carried out by a phosphatase reaction, a terminal transferase reaction, a polymerase reaction, or an exonuclease reaction, or is carried out chemically.

12. Method according to one of Claims 1 through 11, wherein the masking of a 5' end is carried out by removing the phosphate group(s) or by introducing a 5'-modified nucleotide.

13. Method according to one of Claims 1 through 11, wherein a masked 3' end is characterized by the presence of an amino block, a dideoxynucleotide, a 3'-phosphate, or a synthetic 5' end.

14. Method according to one of Claims 1 through 13, wherein the additional nucleic acid molecule forms a hairpin loop, after attachment and/or linkage to the end removed from the prepared nucleic acid molecule, which is used as a primer for the polymerase activity.

15. Method according to one of Claims 1 through 14, wherein the cleavage takes place in step (4) at a predetermined location, using a triple helical DNA which cleaves in a sequence-specific manner, or using a type IIS restriction endonuclease.

16. Method according to Claim 15, wherein the type IIS restriction endonuclease is the RleAI enzyme from *Rhizobium leguminosarum*.

17. Method according to one of Claims 1 through 16, wherein the prepared nucleic acid molecule and/or the additional nucleic acid molecule is/are of synthetic or semisynthetic origin.

18. Method according to one of Claims 1 through 17, wherein the synthesis is at least partially automated.

19, Method according to one of Claims 1 through 18, wherein the synthesis is carried out bound to a matrix.

20. Method according to Claim 19, wherein the matrix is made of Nylon, glass such as CPG and glass

wool, silicate, latex, polystyrene, epoxy, or silicon.

21. Method according to one of Claims 1 through 20, wherein the synthesized nucleic acid molecule is isolated after the synthesis.

22. Method according to one of Claims 1 through 21, wherein after the synthesis with terminal introduction of primer sequences, the synthesized nucleic acid molecule is amplified using a suitable method.

23. Method according to one of Claims 1 through 22, wherein after the synthesis, individual synthesized nucleic acid molecules are connected to form even larger DNA molecular units, optionally mediated by type IIS restriction enzyme termini.

24. Kit comprising:
    (a) a ligase, and/or
    (b) a polymerase,
    (c) optionally a type IIS restriction enzyme,
    (d) optionally a uracil DNA gycosylase and an apyrimidase, and/or an endonuclease III and a formamidopyrimidine DNA glycosylase, and/or a "mismatch repair" enzyme,
    (e) optionally a phosphatase, a terminal transferase, and/or an exonuclease,
    (f) optionally a wash buffer for eluting reaction by-products and material not introduced into the product of the synthesis according to the invention,
    (g) optionally a synthesis matrix, using a nucleic acid molecule which is optionally already bound thereto as starter molecule,
    (h) optionally suitable reaction buffers for the enzymes listed in (a) through (e).

Figure 1

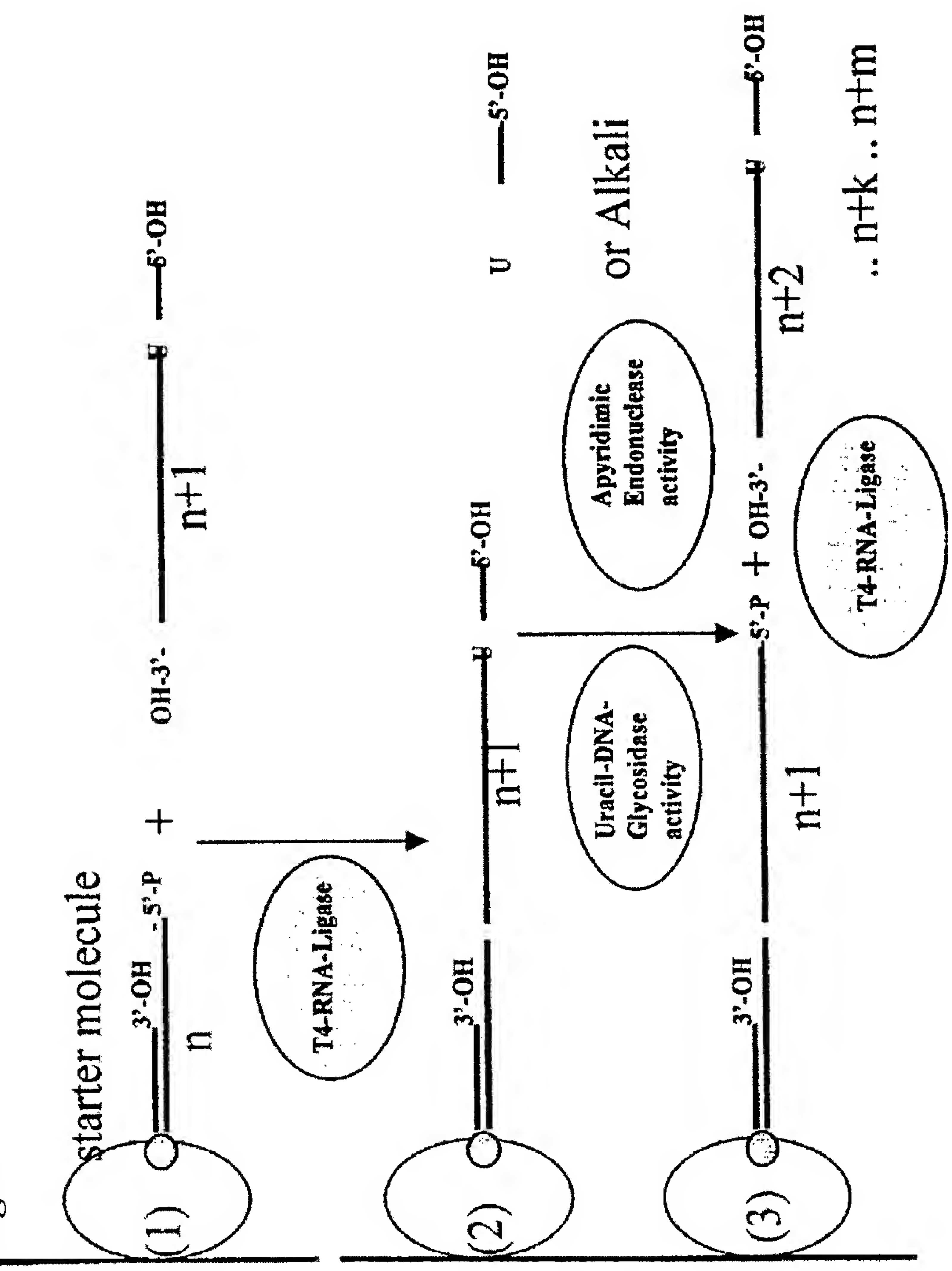

starter molecule
(1)
3'-OH
5'-P
n
+
OH-3'-
n+1
U
5'-OH
T4-RNA-Ligase
(2)
3'-OH
n+1
U
5'-OH
Uracil-DNA-Glycosidase activity
Apyridimic Endonuclease activity
or Alkali
U
5'-OH
(3)
3'-OH
n+1
5'-P
+ OH-3'-
n+2
U
5'-OH
T4-RNA-Ligase
.. n+k .. n+m

Figure 2

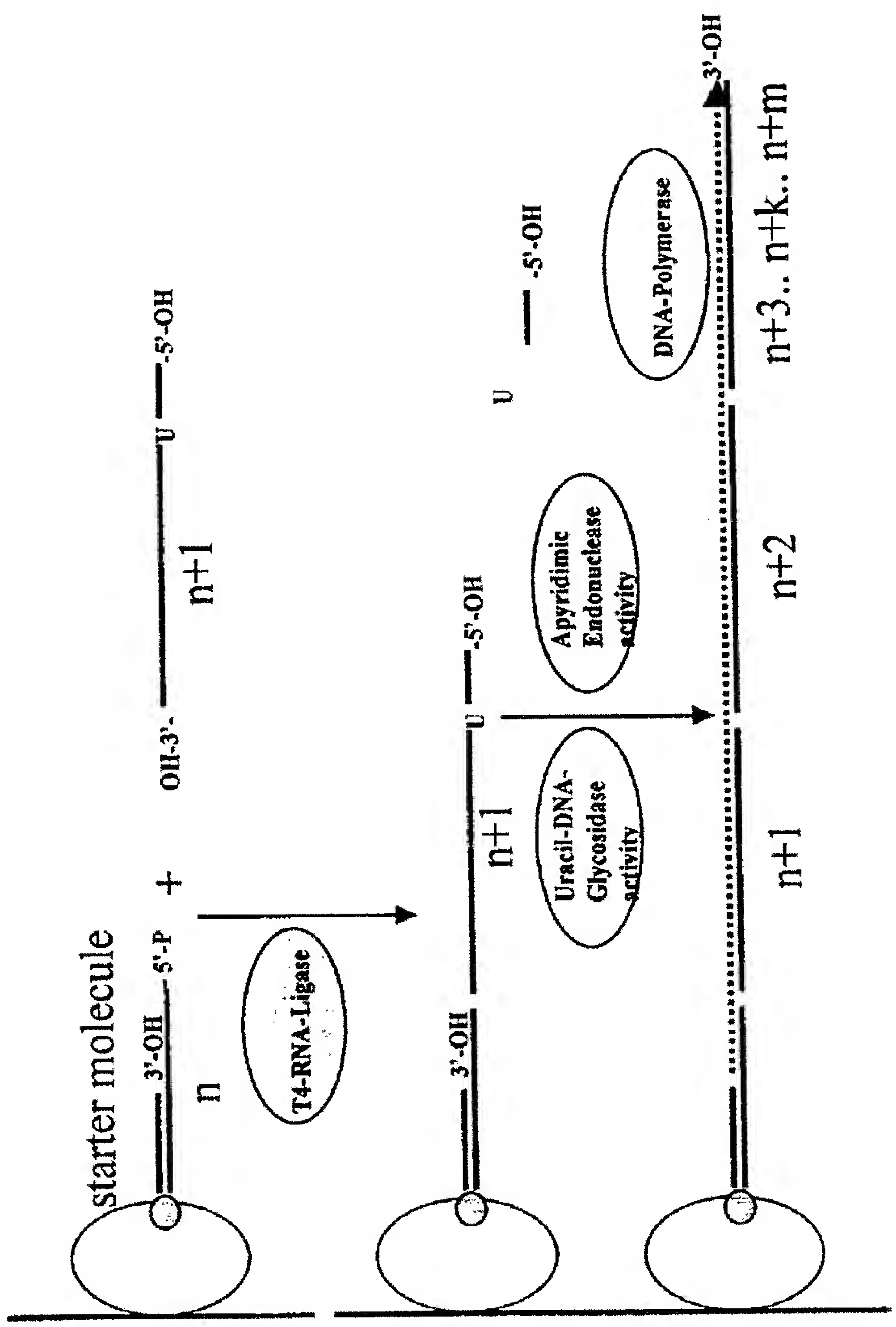

starter molecule
3'-OH
5'-P
n
+
OH-3'-
n+1
U
-5'-OH
T4-RNA-Ligase
3'-OH
n+1
U
-5'-OH
Uracil-DNA-Glycosidase activity
Apyridimic Endonuclease activity
U
-5'-OH
DNA-Polymerase
3'-OH
n+1
n+2
n+3.. n+k.. n+m

Figure 3

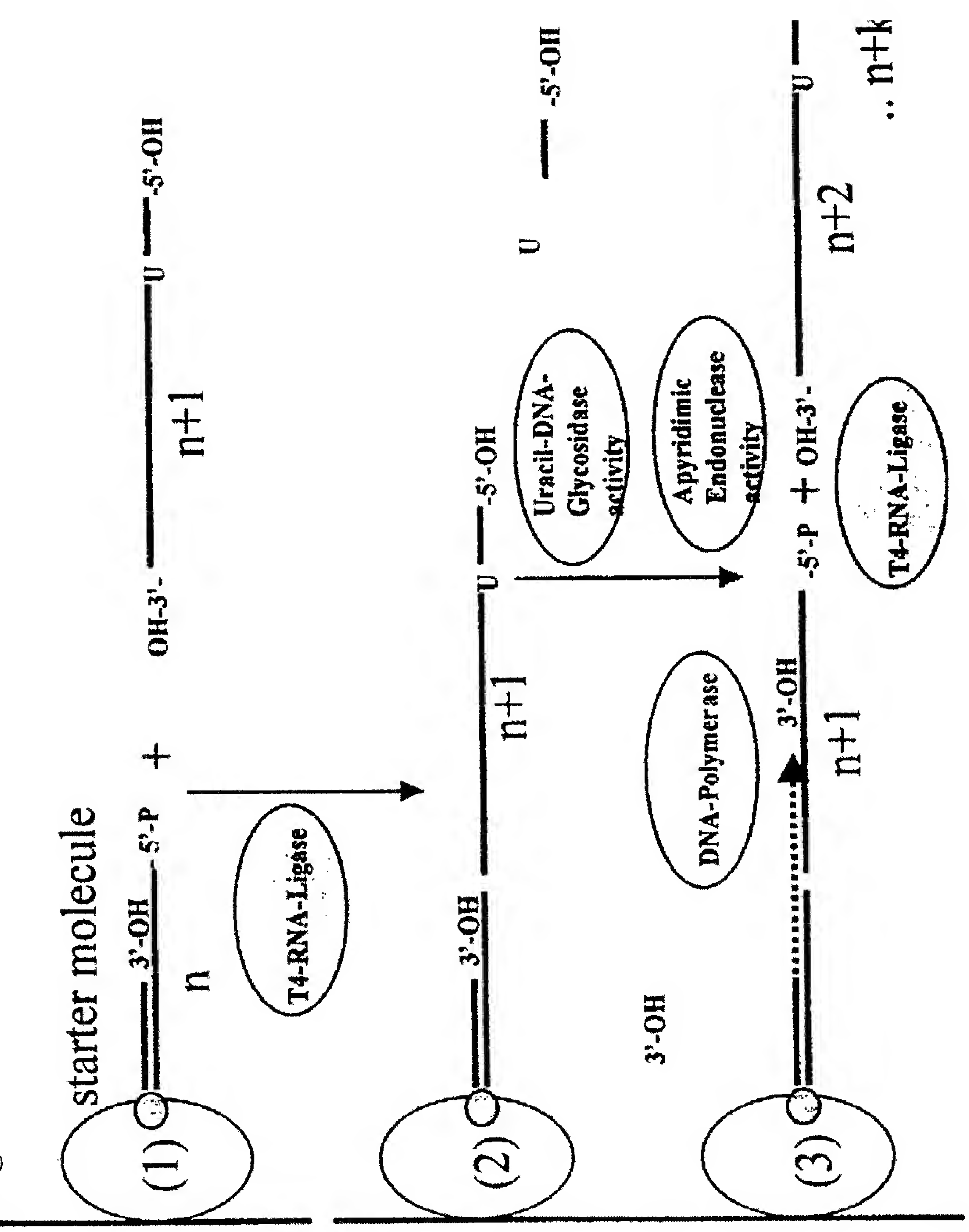
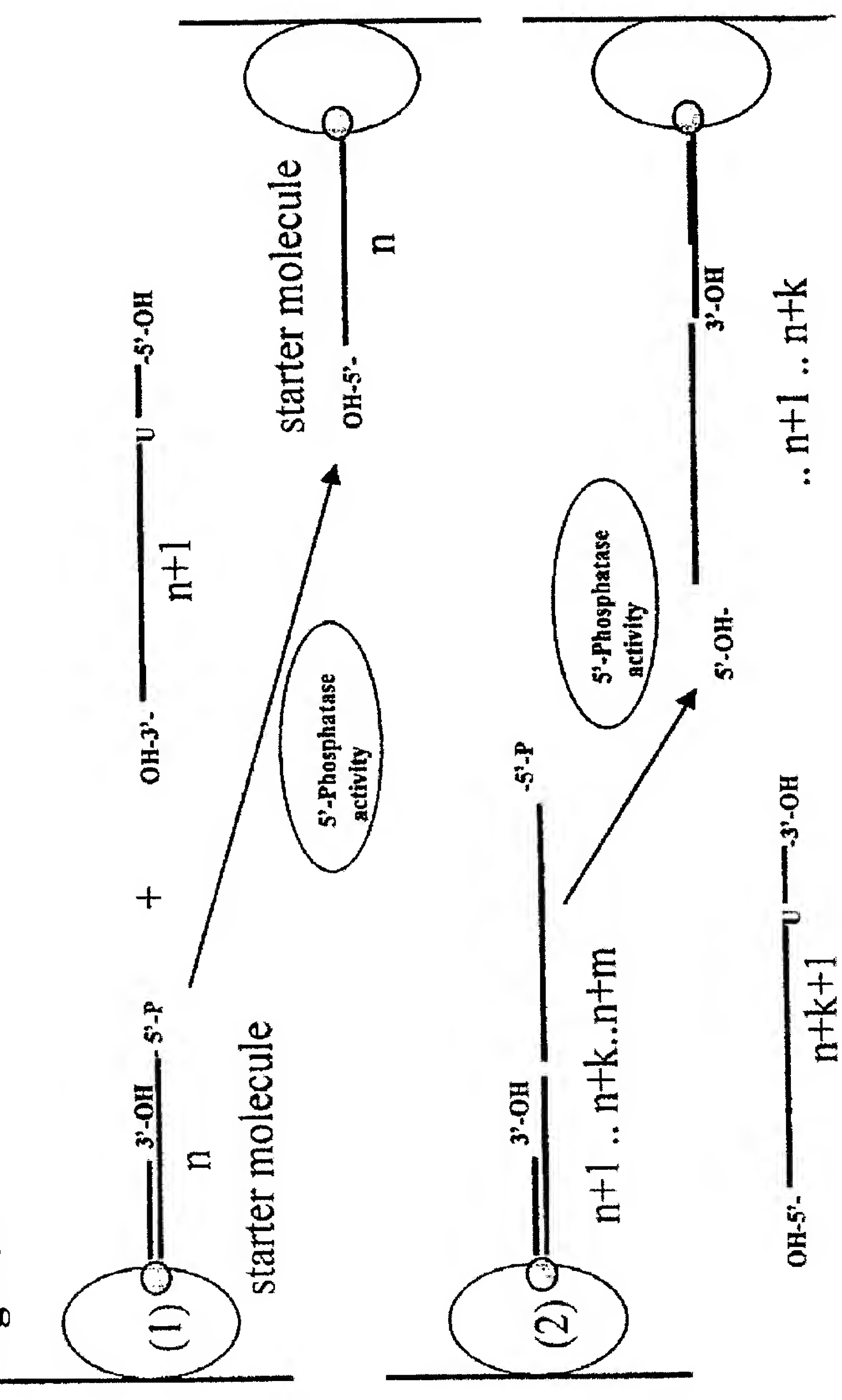

starter molecule
(1)
3'-OH
5'-P
n
+
OH-3'-
U
-5'-OH
n+1
T4-RNA-Ligase
(2)
3'-OH
n+1
U
-5'-OH
U
-5'-OH
Uracil-DNA-Glycosidase activity
Apyridimic Endonuclease activity
3'-OH
DNA-Polymerase
(3)
3'-OH
n+1
-5'-P
+ OH-3'-
T4-RNA-Ligase
n+2
U
.. n+k

Figure 4



(1)
3'-OH
5'-P
n
starter molecule
+
OH-3'-
U
-5'-OH
n+1
5'-Phosphatase activity
starter molecule
OH-5'-
n
(2)
3'-OH
-5'-P
n+1 .. n+k..n+m
OH-5'-
U
-3'-OH
n+k+1
5'-Phosphatase activity
5'-OH-
3'-OH
.. n+1 .. n+k

Figure 5

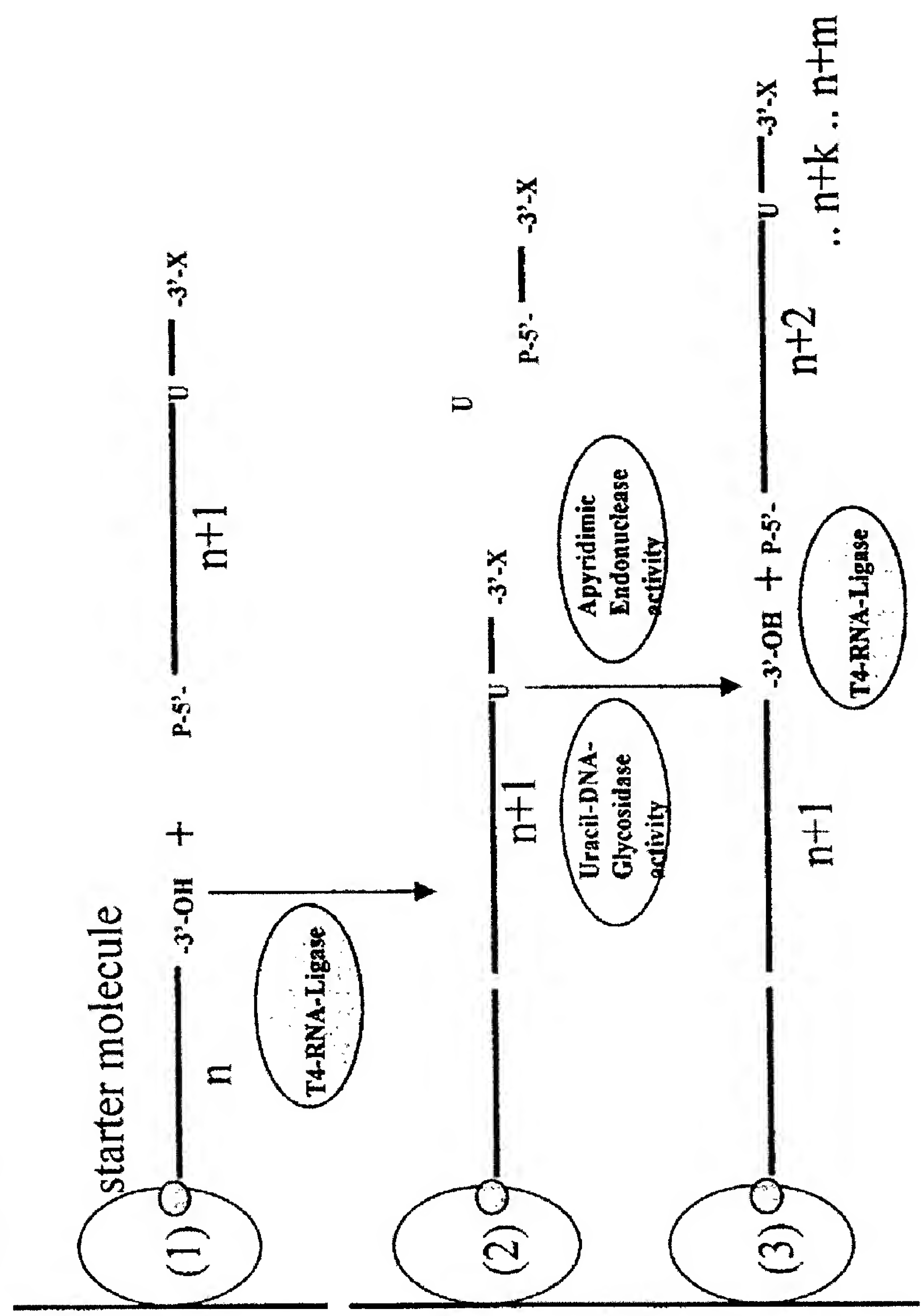
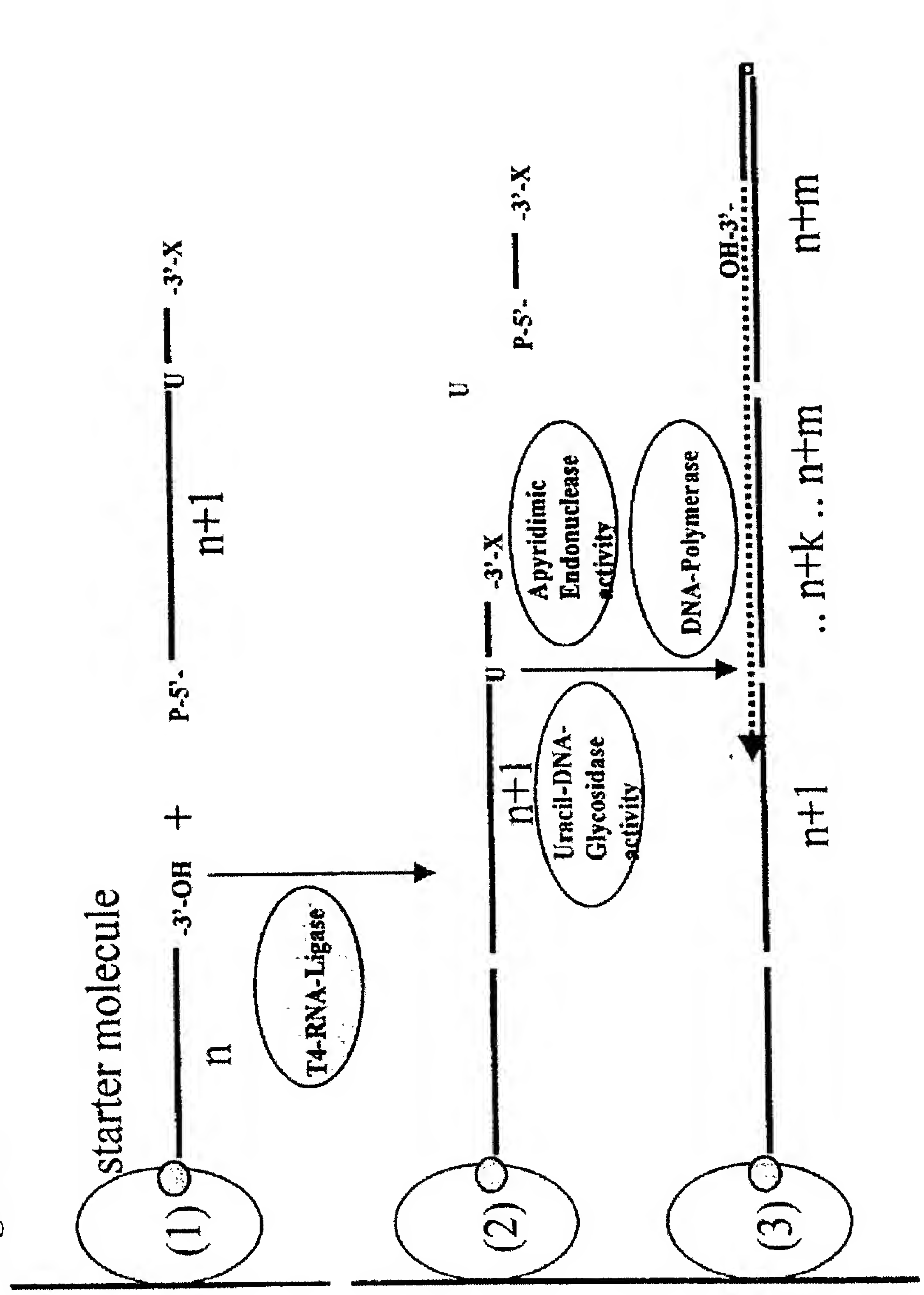

starter molecule
(1)
n
-3'-OH
+
P-5'-
n+1
U
-3'-X
T4-RNA-Ligase
(2)
n+1
U
-3'-X
Uracil-DNA-Glycosidase activity
Apyridimic Endonuclease activity
U
P-5'-
-3'-X
(3)
n+1
-3'-OH
+
P-5'-
n+2
U
-3'-X
T4-RNA-Ligase
.. n+k .. n+m

Figure 6



starter molecule
(1)
n
-3'-OH
+
P-5'-
n+1
U
-3'-X
T4-RNA-Ligase
(2)
n+1
U
-3'-X
U
P-5'-
-3'-X
Uracil-DNA-Glycosidase activity
Apyridimic Endonuclease activity
DNA-Polymerase
(3)
n+1
.. n+k .. n+m
OH-3'-
n+m

Figure 7

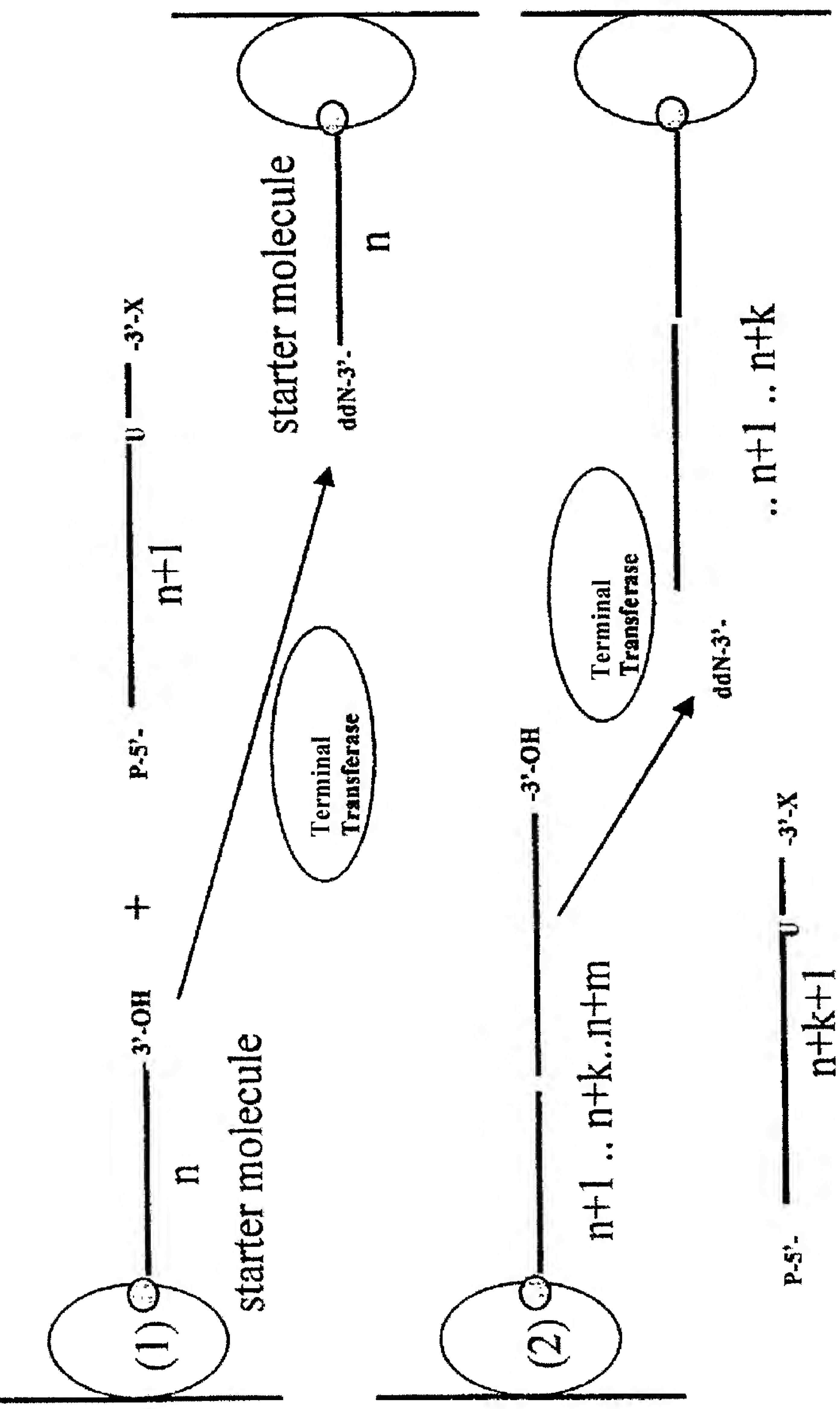

(1)
starter molecule
n
3'-OH
+
P-5'-
n+1
U
-3'-X
Terminal Transferase
ddN-3'-
starter molecule
n
(2)
n+1 .. n+k..n+m
-3'-OH
Terminal Transferase
ddN-3'-
.. n+1 .. n+k
P-5'-
n+k+1
U
-3'-X

Figure 8A

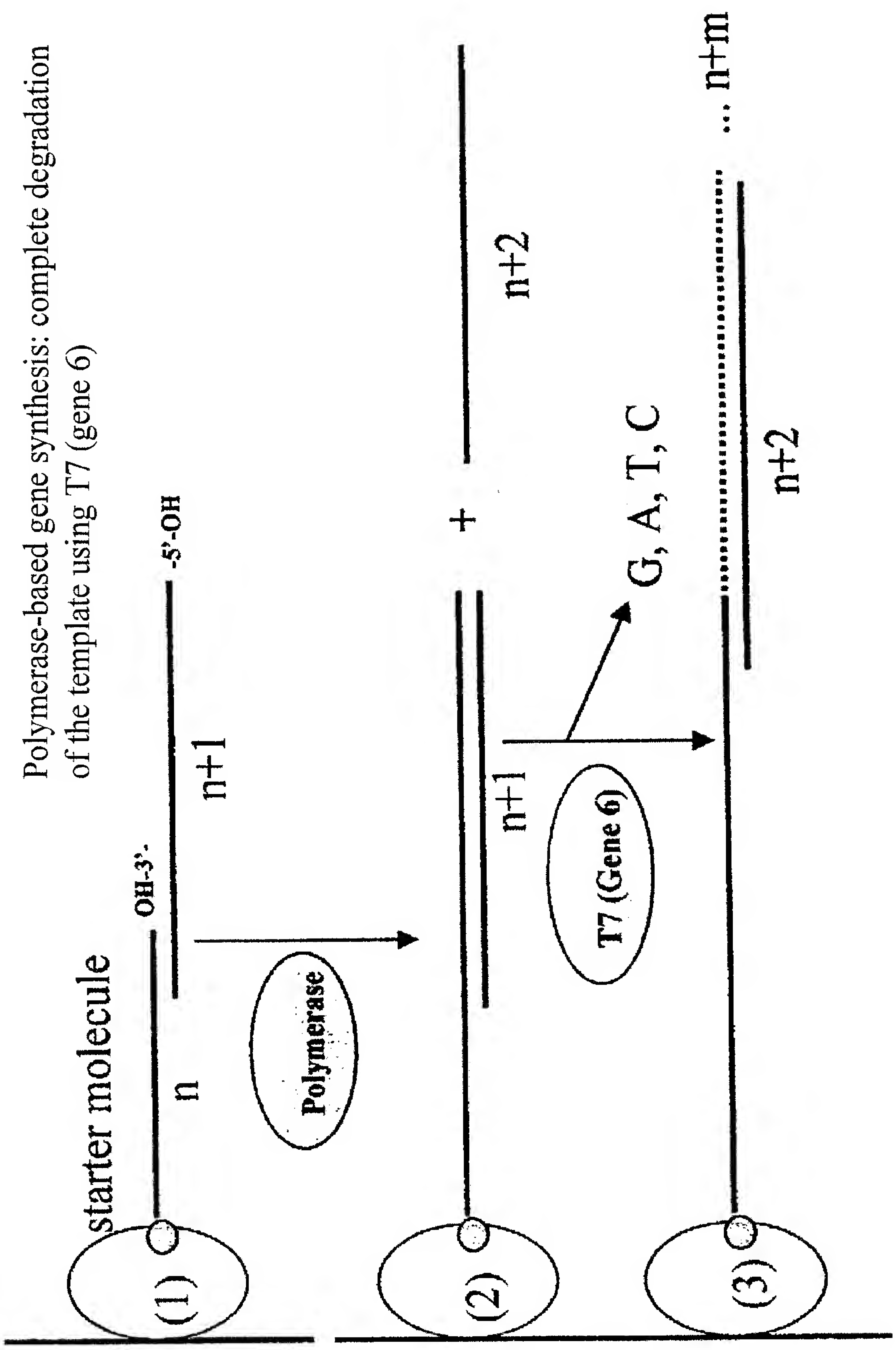

Polymerase-based gene synthesis: complete degradation of the template using T7 (gene 6)
starter molecule
OH-3'-
-5'-OH
n
n+1
Polymerase
(1)
+
n+2
n+1
T7 (Gene 6)
G, A, T, C
(2)
... n+m
n+2
(3)

Figure 8B

Polymerase-based gene synthesis: partial degradation of the terminal double strand using exonuclease III